

# FAX

**ATTN.** Atiba O. Fitzpatrick

---

**Fax Number** 1 571 270 6255

**Phone Number** 571 270 5255

**FROM** Volel Emile, Esq.

---

**Fax Number** 512 306 0240

**Phone Number** 512 306 7969

**SUBJECT** Proposed Claim Amendment For Interview

---

**Number of Pages** 4

**Date** 3/26/2008

**MESSAGE**

---

Examiner Fitzpatrick:

Pleas find enclosed the agenda for this afternoon telephone interview.

Thank you.

Volel

### **BRIEF DESCRIPTION OF THE INVENTION**

Due to recent trends toward telecommuting, mobile offices, and the globalization of businesses, more and more employees are being geographically separated from each other. As a result, less and less face-to-face communications are occurring at the workplace.

Face-to-face communications provide a variety of visual cues that ordinarily help in ascertaining whether a conversation is being understood or even being heard. For example, non-verbal behaviors such as visual attention and head nods during a conversation are indicative of understanding. Certain postures, facial expressions and eye gazes may provide social cues as to a person's emotional state, etc. Non-face-to-face communications are devoid of these cues.

To diminish the impact of non-face-to-face communications, videoconferencing is increasingly being used. A videoconference is a conference between two or more participants at different sites using a computer network to transmit audio and video data. Particularly, at each site there is a video camera, microphone, and speakers mounted on a computer. As participants speak to one another, their voices are carried over the network and delivered to the other's speakers, and the images which appear in front of a video camera appear in a window on the other participant's monitor.

As with any conversation or in any meeting, sometimes a participant might be stimulated by what is being communicated and sometimes the participant might be totally disinterested. Since voice and images are being transmitted digitally, it would be advantageous to store this data to be used later as a speech improving apparatus, system and method.

The present invention provides a method of getting feedback to a public speech to facilitate speech improvement. According to the teachings of the invention, during a speech, data representing audio and video data of participants recorded at a conference where the speech was given is stored. This stored data is used to (1) look for a particular expression exhibited by one of

the participants during the speech, and (2) determine what was being said when the participant exhibited the expression. The analysis of the stored data facilitates may be used for improving speech making by a speaker.

### CLAIM

1. (Currently amended) A speech making improvement method of getting feedback to a public speech to facilitate speech improvement. the method using stored data, the data being audio and video data of participants recorded at a conference where the speech was given, the method comprising the steps of:

indicating an expression for which to search. the expression being that may have been exhibited by one or more of the participants at the conference during the speech conference to search for;

determining, using the stored data in conjunction with an automated facial decoding system, whether at least one participant exhibited the indicated expression; and

analyzing, in response to determining that the at least one participant exhibited the expression, the video data representing the at least one participant exhibiting the expression and the audio data representing what was being said in the speech when the at least one participant exhibited the expression to improve a speaker's speech making ability.

### ARGUMENTS

Erten purports to teach an audio visual speech processing system. According to Erten, the system combines audio signals that register the voice or voices of one or more speakers with video signals that register the image of

faces of these speakers. This results in enhanced speech signals and improved recognition of spoken words.

By contrast the present invention uses audio signals that register the voice of a speaker at a conference with video signals of participants exhibiting a particular expression at the conference for speech improvement. Specifically, a particular expression of a participant during the speech at the conference is used in conjunction with what was being said at the time by a speaker for speech improvement.

Thus, Erten does not teach or show *indicating an expression for which to search, the expression being exhibited by one or more of the participants at the conference during the speech; determining, using the stored data in conjunction with an automated facial decoding system, whether at least one participant exhibited the indicated expression; and analyzing, in response to determining that the at least one participant exhibited the expression, the video data representing the at least one participant exhibiting the expression and the audio data representing what was being said in the speech when the at least one participant exhibited the expression to improve a speaker's speech making ability* as claimed.